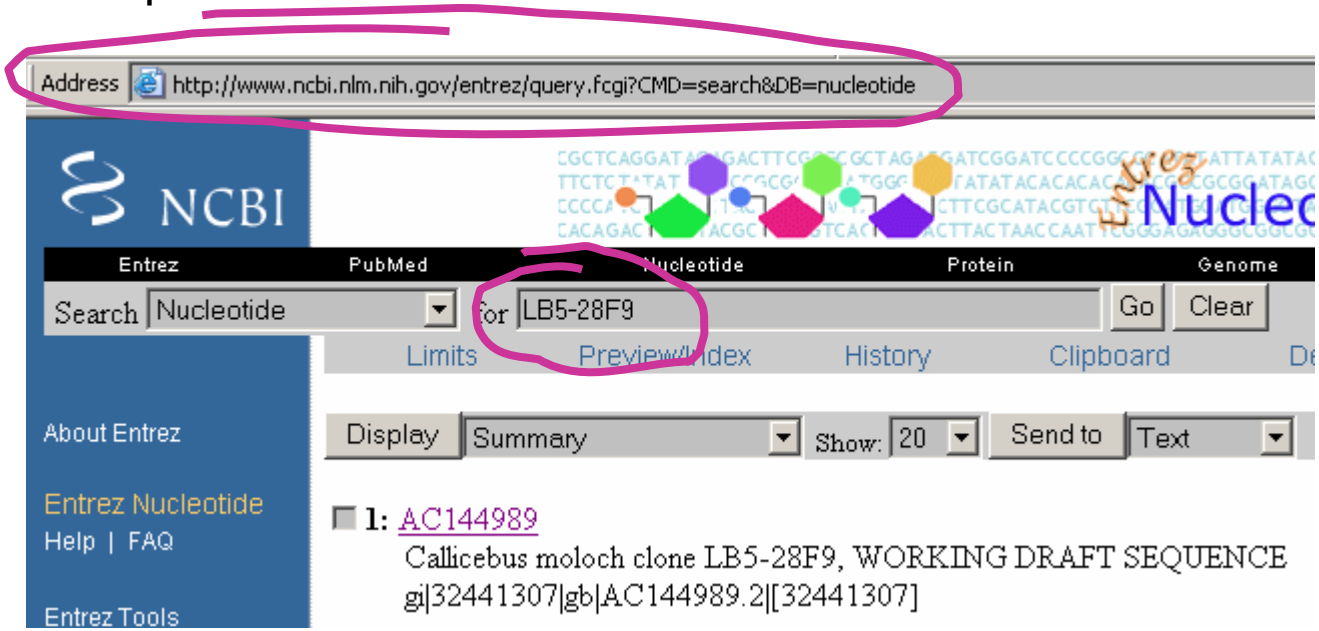
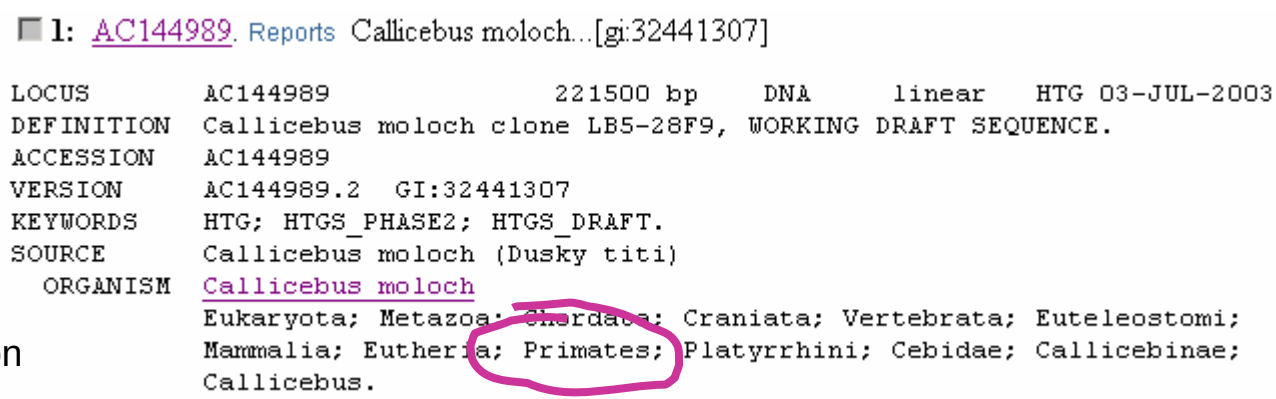
Step-by-step key to Day 1 exercises.

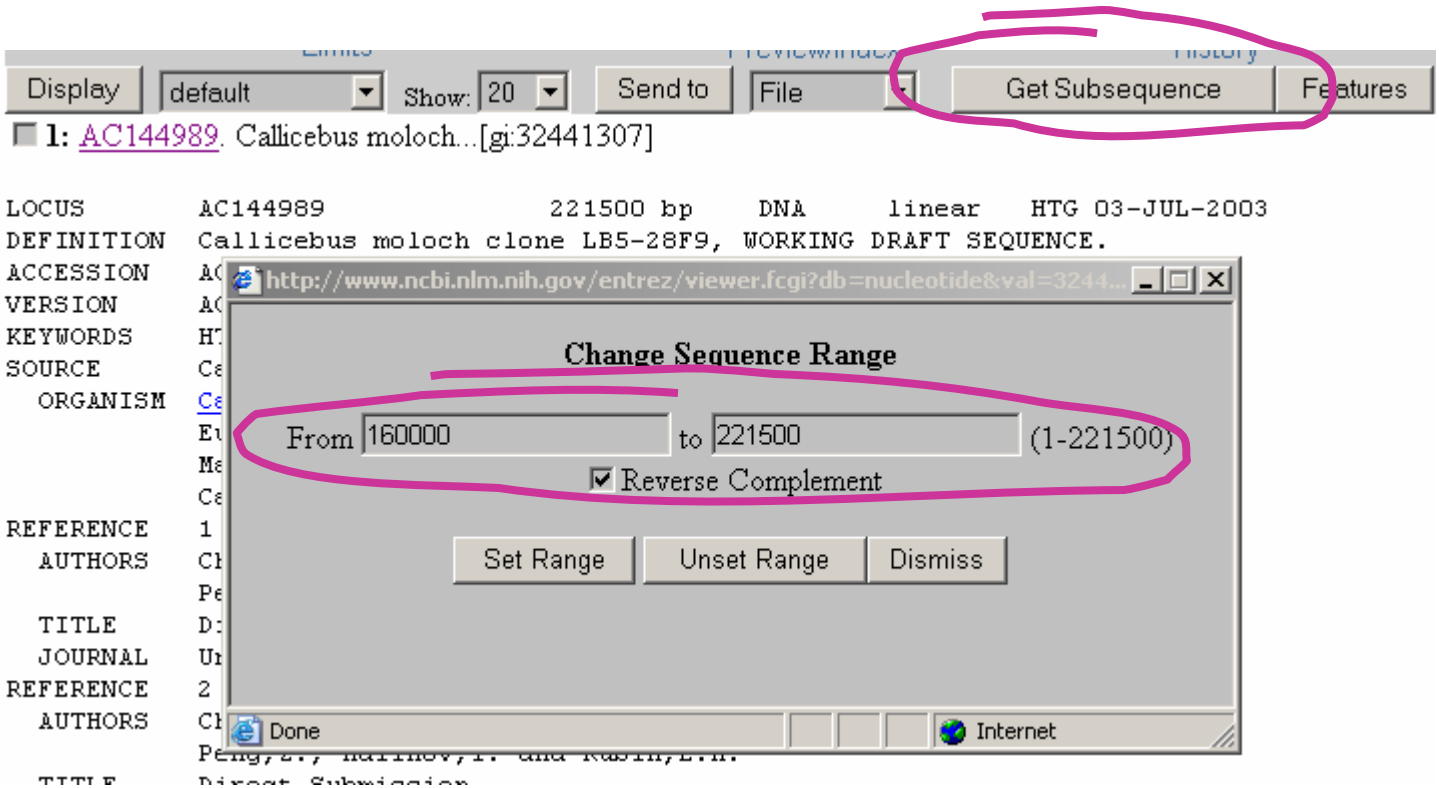
Exercise #1: **Retrieve the sequence of the BAC LB5-28F9 from GenBank**



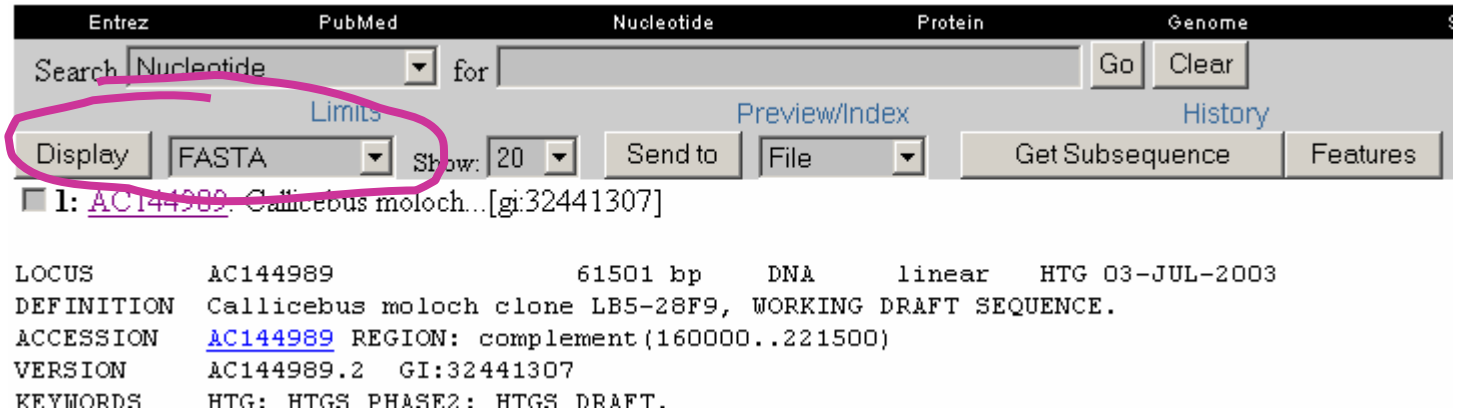


Species information

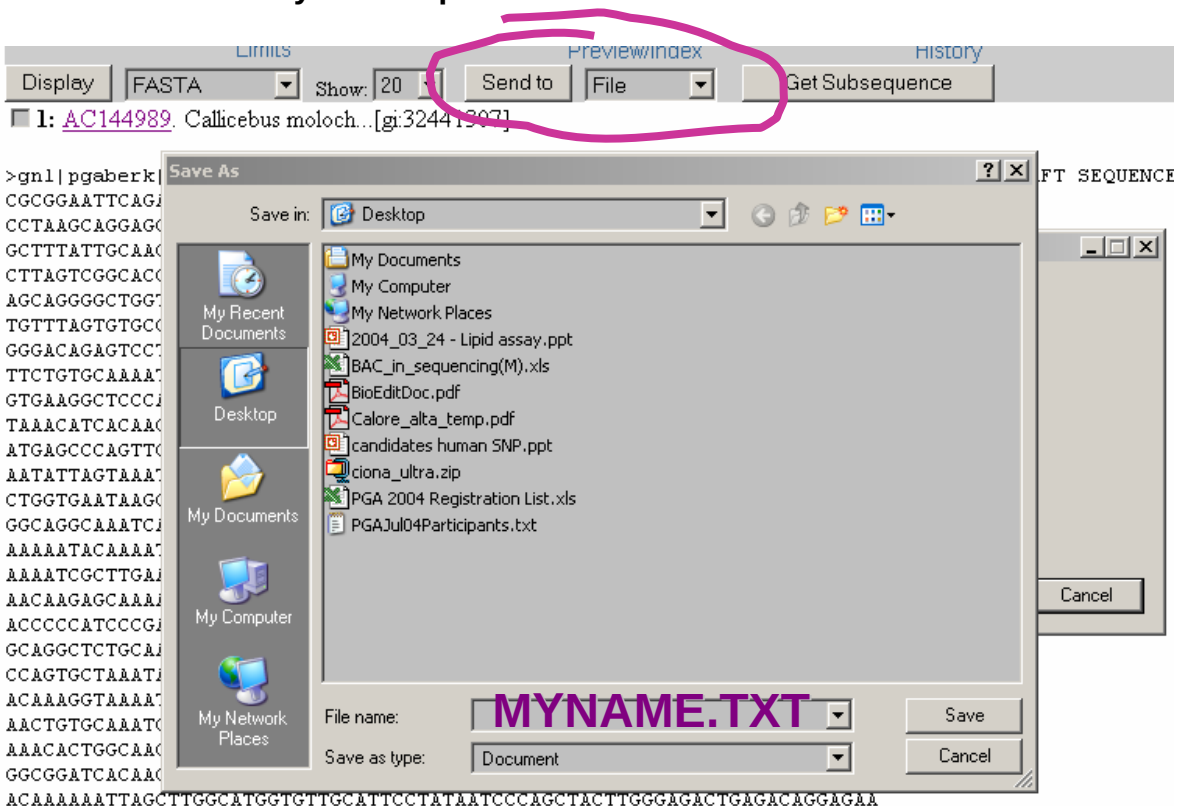
Exercise #2: **Retrieve a subsequence 160000-221500 and reverse complement it**



Exercise #2: **Convert the extracted sequence to FASTA format**



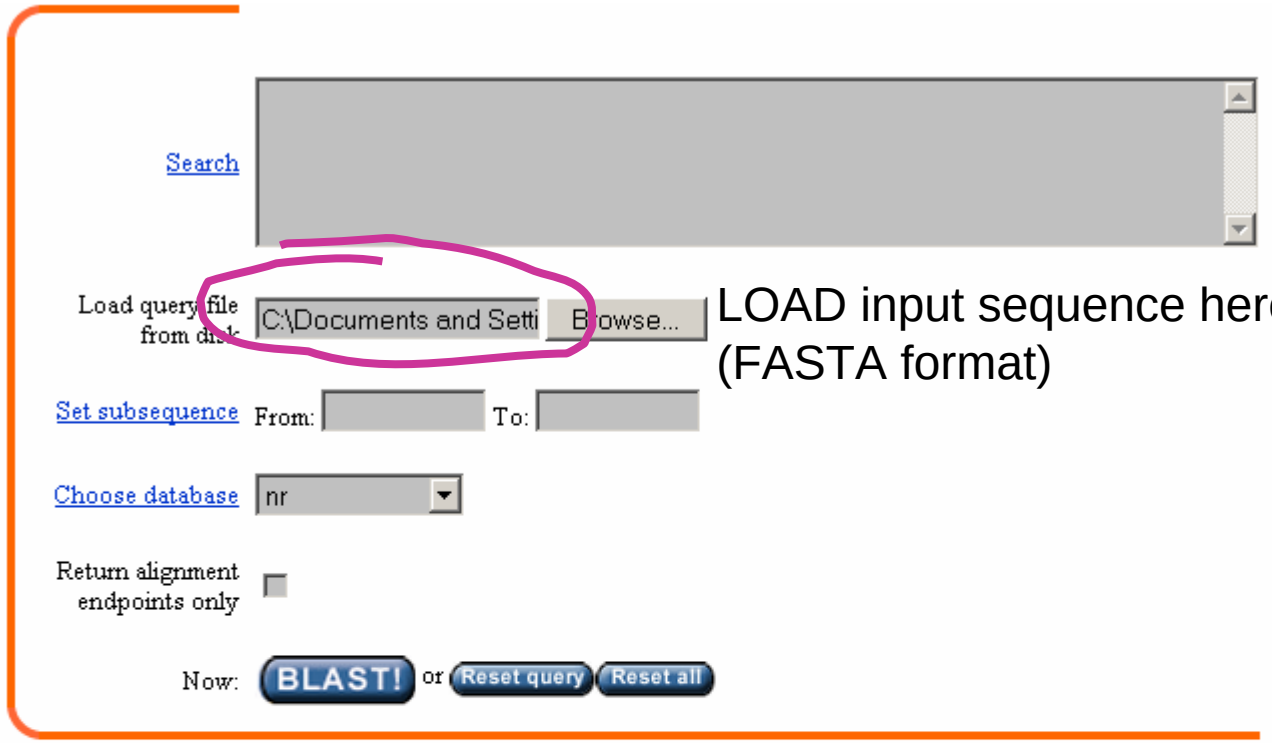
And save as a text document on your computer.



Exercise #3: **Find what coding sequences are present in your extracted sequence**

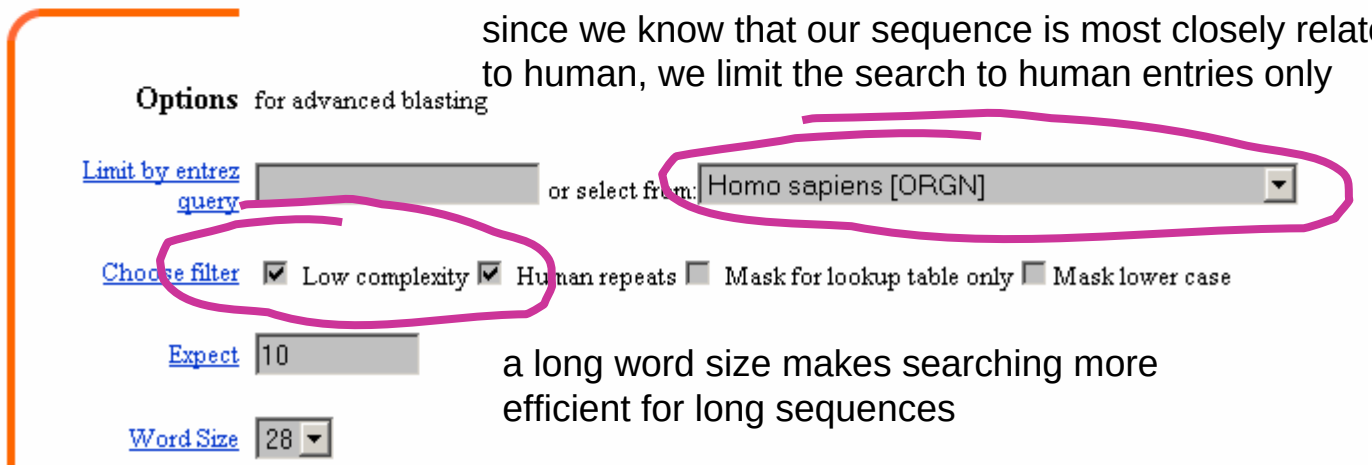
Use Megablast

Exercise #3: How to run Megablast



LOAD input sequence here (FASTA format)

since we know that our sequence is most closely related to human, we limit the search to human entries only



a long word size makes searching more efficient for long sequences

Exercise #4: BLAST Results for BAC sequence

Query: your BAC sequence - **Subject:** database hit

NCBI *results of* **BLAST**

e-value: significance of the alignment. Alignments with the lowest e-value are the most significant. However, remember that long alignments have lower e-values

Subject accession #: look in the NUCLEOTIDE database for information on these sequences (sse next page)

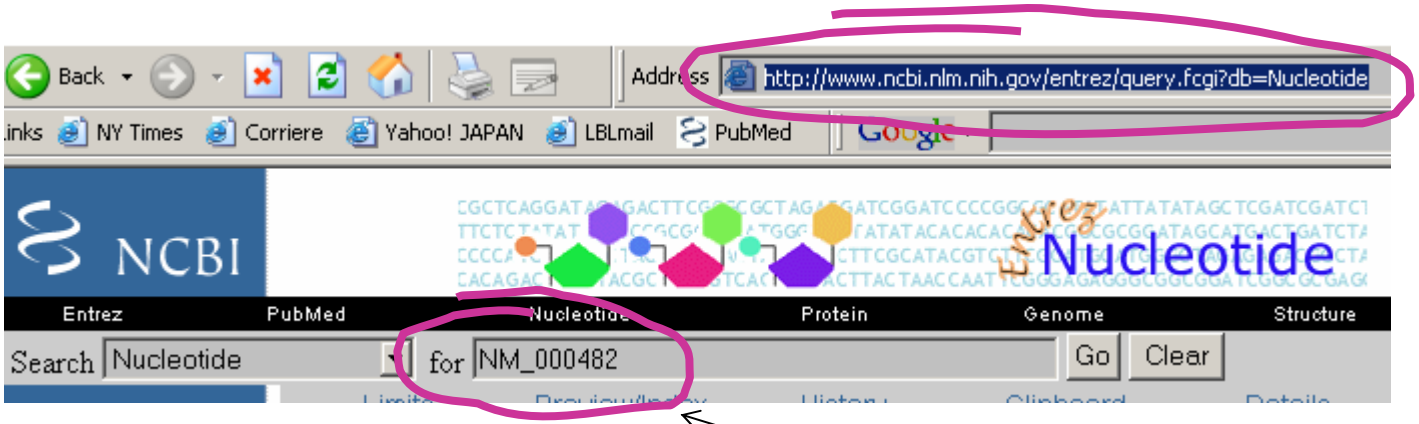
```
# BLASTN 2.2.9 [May-01-2004]
# Query: Callicebus moloch clone L
# Database: nr
# Fields: Query id, Subject id, % identity, alignment length, mismatches, gap openings, q. start, q. end, s. start, s. end, e-value, bit score
Callicebus      gi|9625348|gb|AC074203.3|       89.58   4817    320     110     47821   52528   35822   40565   0.0     6191
Callicebus      gi|9625348|gb|AC074203.3|       90.65   1797    117     29      52562   54324   40588   42379   0.0     2475
Callicebus      gi|9625348|gb|AC074203.3|       88.95   1883    136     37      43350   45192   31651   33501   0.0     2352
Callicebus      gi|9625348|gb|AC074203.3|       88.74   1235    84      33      46617   47823   34590   35797   0.0     1519


Callicebus      gi|178756|gb|J02758.1|HUMAPOA4A 89.22   306     20      6       20149   20450   3307    3607    3e-105  396
Callicebus      gi|22091457|ref|NM_052968.2|    93.28   982     60      5       45271   46247   371     1351    0.0     1502
Callicebus      gi|22091457|ref|NM_052968.2|    90.66   503     29      10      46617   47117   1381    1867    0.0     679
Callicebus      gi|22091457|ref|NM_052968.2|    96.55   116     4       0       44451   44566   55      170     3e-46   200
Callicebus      gi|22091457|ref|NM_052968.2|    95.00   120     6       0       45073   45192   173     292     5e-45   196
Callicebus      gi|6707432|gb|AF202889.1|       93.28   982     60      5       45271   46247   371     1351    0.0     1502
```

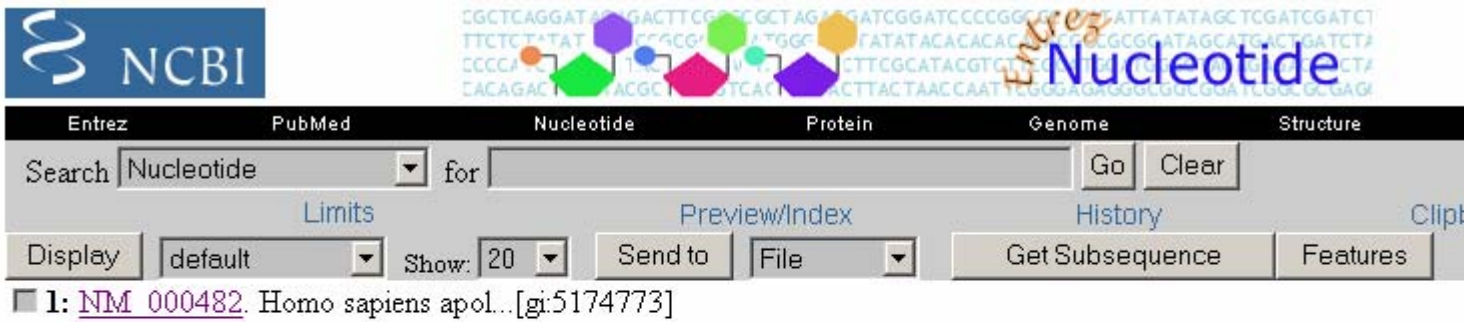
3 RefSeq hits: this is a coding sequence (mRNA)

```
Callicebus      gi|49902035|gb|BC074764.1|      93.26   89      6       0       17677   17765   1282    1194    4e-27   137
Callicebus      gi|5174773|ref|NM_000482.2|     91.94   707     53      4       19254   19958   522     1226    0.0     1027
Callicebus      gi|5174773|ref|NM_000482.2|     92.73   165     9       3       17603   17765   1       164     2e-59   244
Callicebus      gi|5174773|ref|NM_000482.2|     96.88   128     4       0       18124   18251   163     290     4e-53   223
Callicebus      gi|563319|gb|M10373.1|HUMAPOAIVA        91.95   708     51      6       19254   19958   6       710     0.0     1027
Callicebus      gi|178778|gb|M14566.1|HUMAPOAIV 91.94   707     53      4       19254   19958   348     1052    0.0     1027
Callicebus      gi|178778|gb|M14566.1|HUMAPOAIV 86.98   192     11      4       20149   20337   1108    1292    4e-53   223
Callicebus      gi|178778|gb|M14566.1|HUMAPOAIV 97.41   116     3       0       18136   18251   1       116     6e-48   206
Callicebus      gi|178758|gb|M13654.1|HUMAPOA4B 91.95   708     51      6       19254   19958   459     1163    0.0     1027
Callicebus      gi|178758|gb|M13654.1|HUMAPOA4B 96.88   128     4       0       18124   18251   100     227     4e-53   223
Callicebus      gi|178758|gb|M13654.1|HUMAPOA4B 95.05   101     5       0       17665   17765   1       101     9e-36   166
Callicebus      gi|28771|emb|X02162.1|HSAPOAIB  90.83   665     53      8       5427    6087    287     947     0.0     919
Callicebus      gi|28771|emb|X02162.1|HSAPOAIB  91.03   156     5       2       4688    4842    129     283     2e-55   231
Callicebus      gi|45573201|ref|NM_000039.1|    90.80   663     53      8       5427    6085    239     897     0.0     915
Callicebus      gi|45573201|ref|NM_000039.1|    91.03   156     5       2       4688    4842    81      235     2e-55   231
Callicebus      gi|178772|gb|M27875.1|HUMAPOAIB 90.80   663     53      8       5427    6085    220     878     0.0     915
```

<u>Exercise #4:</u> Find out what are the genes present in your BAC sequence by lookin in the "nucleotide" database



search the nucleotide database for the accession # found in your BLAST search



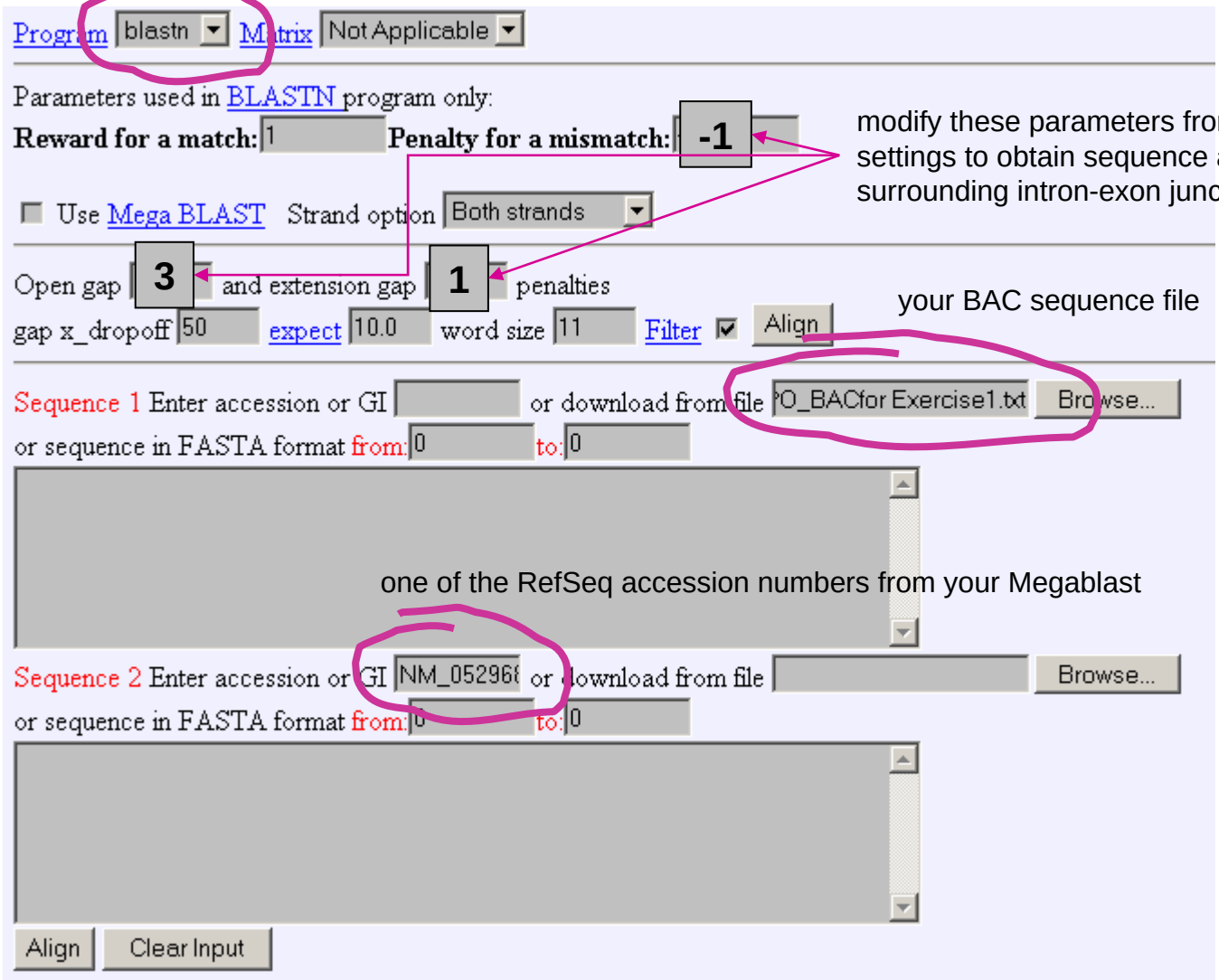

```
LOCUS       NM_000482                1466 bp    mRNA    linear   PRI 20-DEC-2003
DEFINITION  Homo sapiens apolipoprotein A-IV (APOA4), mRNA.
ACCESSION   NM_000482
VERSION     NM_000482.2  GI:5174773
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
```

<u>Exercise #5:</u> Use Bl2seq to locate the intron-exon boundaries of your predicted coding sequences

(TIP: work with multiple browser windows to copy and paste information)

Exercise #5: Compare your BAC DNA sequence with the sequence of the human APOA5 cDNA

blastn is the program to compare DNA to DNA



modify these parameters from the default settings to obtain sequence alignments surrounding intron-exon junctions

your BAC sequence file

one of the RefSeq accession numbers from your Megablast

Exercise #5: Use Bl2seq to locate the intron-exon boundaries of your predicted coding sequences

Intron-Exon junction rule: exon1 **GT**·····**AG** exon2

```
BLAST 2 SEQUENCES RESULTS VERSION BLASTN 2.2.10
Sequence 1  gnl|pgaberk|T022-28F9:c221500-160000 Callicebus moloch clone LB5-28F9, WORKING DRAFT
SEQUENCE Length 61501 (1 .. 61501)
Sequence 2   gi 22091457 Homo sapiens apolipoprotein A-V (APOA5), mRNA Length 1889 (1 .. 1889)
Intron sequences are in bold and red. Look for consensus splice sites: GT-intron-AG (GT and AG are part of
the intron)
Exon 3
Query:            45065 cccaggagcctgaaagacagccttgagcaagacctcaacaatatgaacaagttcctggaa 45124
                        ||| || ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:            165   cccgggaccctgaaagacagccttgagcaagacctcaacaatatgaacaagttcctggaa 224
apolipoprotein AV 50     P  A  T  L  K  D  S  L  E  Q  D  L  N  N  M  N  K  F  L  E
.....................................................................................
Query:            46201 cactcaccaggctttgcaaacccagctttccagtgctcatttgggaattctcataagttg 46260
                        || |||||||||||||||||||| | |||||||||||||||||| ||||| ||||
Sbjct:            1305  cattcaccaggctttgcaaacccagcctcccagtgctcatttgggaatgctcatgagtta 1364

Exon 4
Query:            46601 agctgctcctttcaag-----gggagtagggagggagaaaggcaccatgcatgtgggtga 46655
                        |||| |||| ||||||     |||||||||||||||  ||||||||||||||||||||
Sbjct:            1360  agttactccattcaagggtgagggagtagggagggag--aggcaccatgcatgtgggtga 1417
.....................................................................................
Query:            47075 gaagcctagacttctggctcaaatgaattagatgtttatgatagaa 47120
                        |||||||||||||||||||||||||||| ||||||||||||||| ||
Sbjct:            1825  gaagcctagacttctggctcaaatgaaatagatgtttatgataaaa 1870

Exon 2
Query:            44451 cagcgttttcggccacccaggcacggaaaggcttctgggactacttccgccagaccagcg 44510
                        |||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Sbjct:            55    cagcgttttcggccacccaggcacggaaaggcttctgggactacttcagccagaccagcg 114
apolipoprotein AV 13    S  A  F  S  A  T  Q  A  R  K  G  F  W  D  Y  F  S  Q  T  S

Query:            44511 gggacaaaggcaggatggagcagatccatcagcagaagatggctcgtgaacccgcgg 44566
                        |||||||||||||| ||||||||||||||||||||||||||||| || ||||||
Sbjct:            115   gggacaaaggcagggtggagcagatccatcagcagaagatggctcgcgagcccgcg 170
apolipoprotein AV 33    G  D  K  G  R  V  E  Q  I  H  Q  Q  K  M  A  R  E  P  A

Exon 1
Query:            44265 agcagacaatggcaagcatggctgtcgtgctcacctgggctctggctctcctctcag 44321
                        |||||  |||||||||||||||| |||||||||||||||||||||| || ||||
Sbjct:            1     agcaggtaatggcaagcatggctgccgtgctcacctgggctctggctcttctttcag 57
apolipoprotein AV 1                       M  A  A  V  L  T  W  A  L  A  L  L  S
```

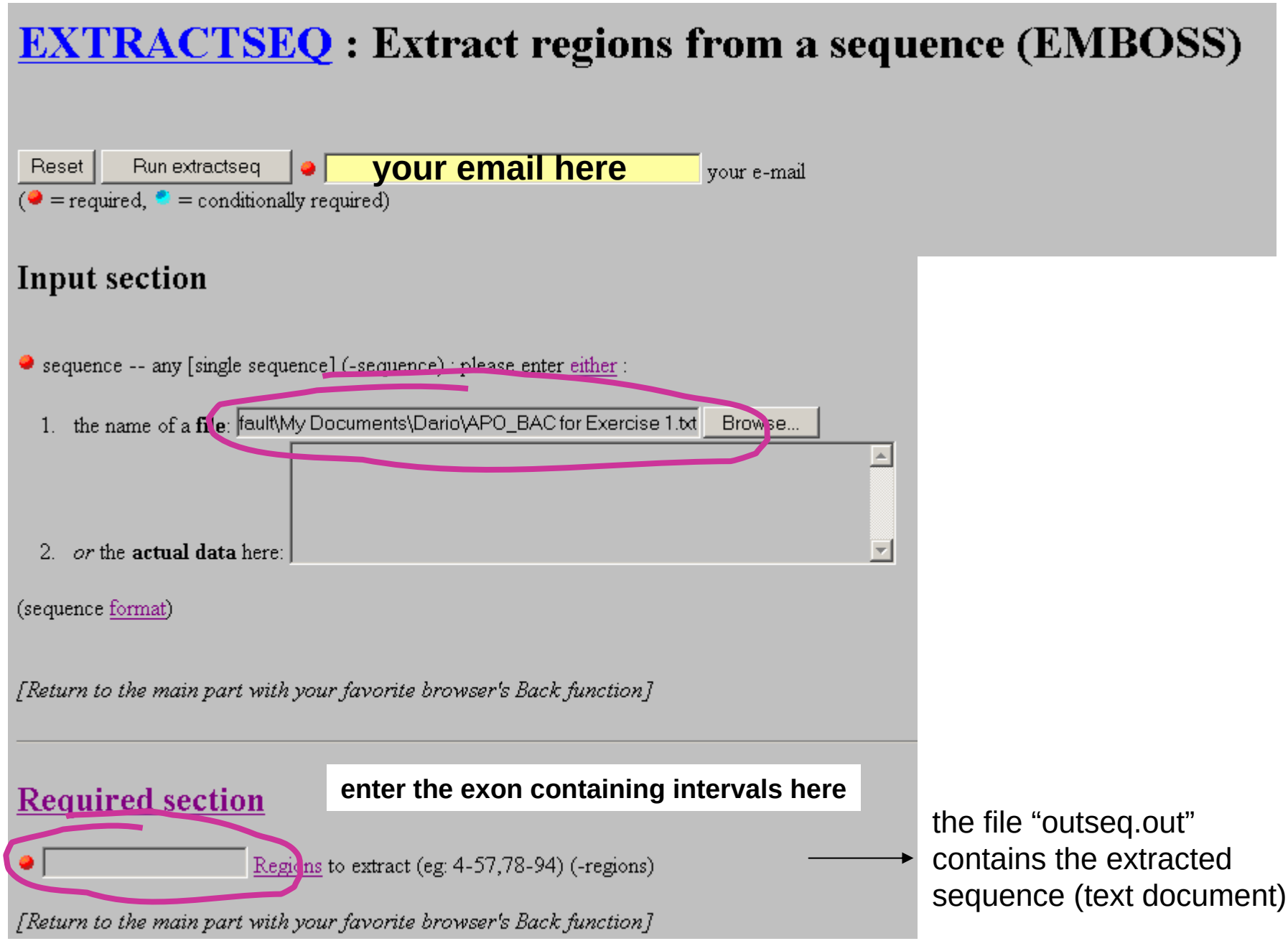
your BAC sequence

human APOA5 cDNA

human APOA5 protein sequence

Exercise #6: **Use EXTRACTSEQ to extract the sequence of your predicted C. moloch APOA5 cDNA**

bioweb.pasteur.fr/seqanal/interfaces/extractseq.html



the file “outseq.out” contains the extracted sequence (text document)

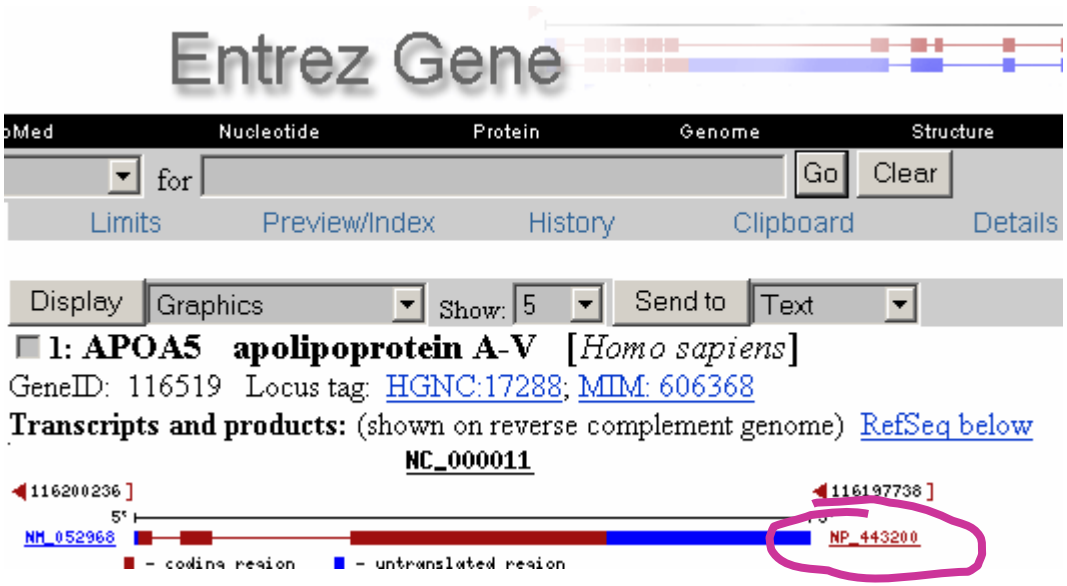
Exercise #6: **UseGETORF on your EXTRACTSEQ output to verify that the cDNA you extracted contains an intact open reading frame**

Exercise #7: Retrieve the human APOA5 protein sequence . There are several ways to do this:

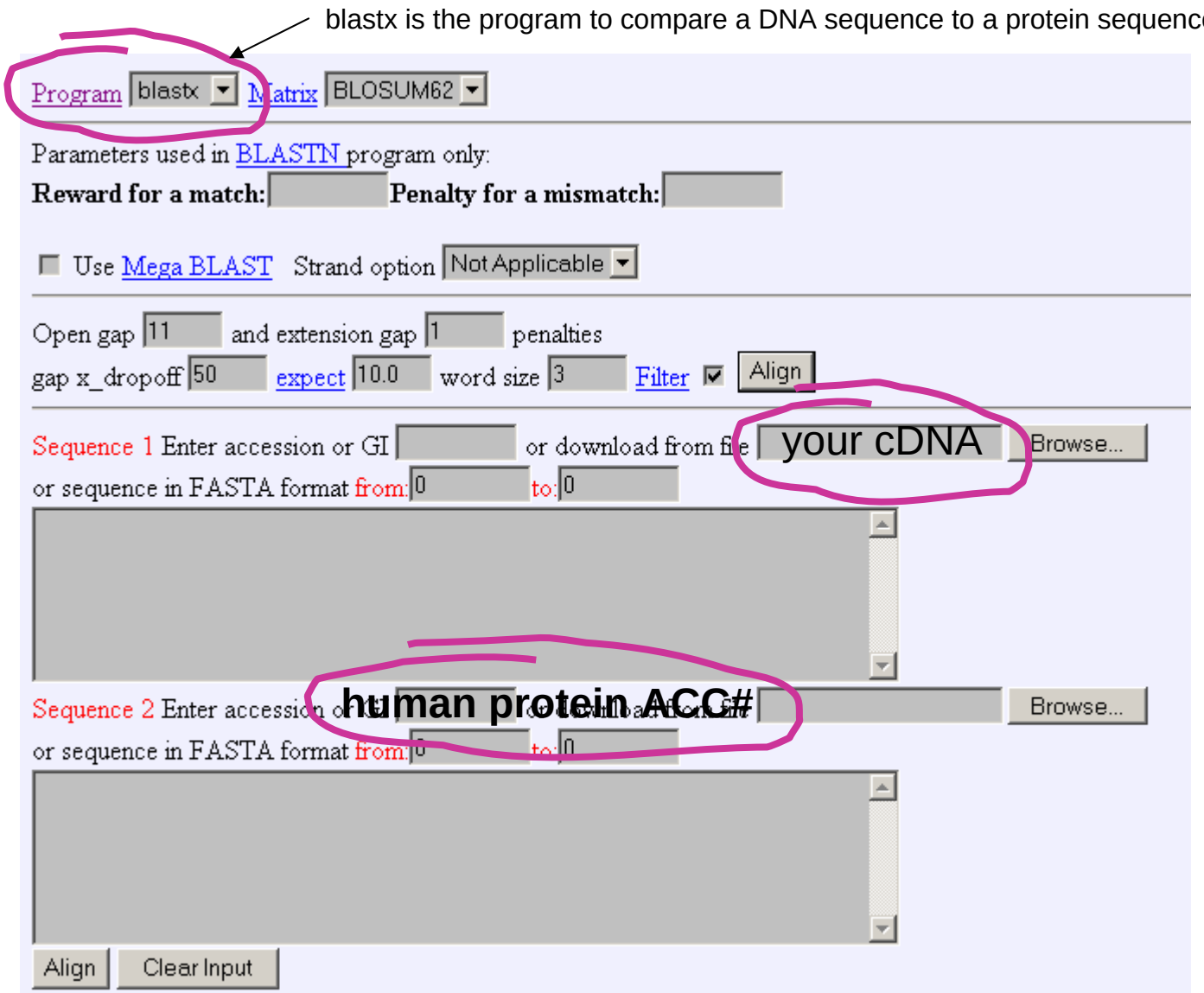
1. You can find that information in the GenBank (nucleotide database) page for the mRNA accession #: look for RefSeq product, this is the protein sequence accession number

```
     CDS             18..1109
                     /gene="APOA5"
                     /note="regeneration-associated protein 3; apolipoprotein
                     A5;
                     go_function: lipid binding [goid 0008289] [evidence IEA];
                     go_process: lipid transport [goid 0006869] [evidence IEA]"
                     /codon_start=1
                     /product="apolipoprotein AV"
                     /protein_id="NP_443200.1"
                     /db_xref="GI:16445025"
                     /db_xref="GeneID:116519"
                     /db_xref="LocusID:116519"
                     /db_xref="MIM:606368"
                     /translation="MAAVLTWALALLSAFSATQARKGFWDYFSQTSGDKGRVEQIHQQ
```

2. You can query the ENTREZ gene database with the mRNA accession # (http://www.ncbi.nlm.nih.gov/gquery/gquery.fcgi). Look for results in the “Gene” section.
   Open the gene page and for the NP_xxxxxx accession number



Exercise #7: compare the APOA5 cDNA sequence of your BAC and of human protein:



Exercise #7: **Blastx alignment of Callicebus and human APOA5**

```
Score =  676 bits (1745), Expect = 0.0
Identities = 340/363 (93%), Positives = 352/363 (96%)
 Frame = +3
```

```
Query: 18   MAVVLTWALALLSAFSATQARKGFWDYFRQTSGDKGRMEQIHQQKMAREPASLKDSLEQD 197
            MA VLTWALALLSAFSATQARKGFWDYF QTSGDKGR+EQIHQQKMAREPA+LKDSLEQD
Sbjct: 1    MAAVLTWALALLSAFSATQARKGFWDYFSQTSGDKGRVEQIHQQKMAREPATLKDSLEQD 60

Query: 198  LNNMMNKFLERLGPLSGSEAPRIPREPVGMRQQLQEELEEVRARLQPHMAEAHELVGWNLE 377
            LNNMMNKFLE+L PLSGSEAPR+P++PVGMR+QLQEELEEV+ARLQP+MAEAHELVGWNLE
Sbjct: 61   LNNMMNKFLEKLRPLSGSEAPRLPQDPVGMRRQLQEELEEVKARLQPYMAEAHELVGWNLE 120

Query: 378  GLRQQLKPYTMDLMEQVALRVQELQEQLRVVGEDTKAQLLGGVGEARALLQELQSRVVHH 557
            GLRQQLKPYTMDLMEQVALRVQELQEQLRVVGEDTKAQLLGGV EA ALLQ LQSRVVHH
Sbjct: 121  GLRQQLKPYTMDLMEQVALRVQELQEQLRVVGEDTKAQLLGGVDEAWALLQGLQSRVVHH 180

Query: 558  TGRFKELFHPYAESLVSGIGRHVQELHRSVAPHAPASPARLSRCVQVLSRKLTLKAKALH 737
            TGRFKELFHPYAESLVSGIGRHVQELHRSVAPHAPASPARLSRCVQVLSRKLTLKAKALH
Sbjct: 181  TGRFKELFHPYAESLVSGIGRHVQELHRSVAPHAPASPARLSRCVQVLSRKLTLKAKALH 240

Query: 738  ARIQQNLDQLREELSRAFAGTGAEQGAGPDPQMLSEEVRQRLQAFRQDTYLQIAAFTRAI 917
            ARIQQNLDQLREELSRAFAGTG E+GAGPDPQMLSEEVRQRLQAFRQDTYLQIAAFTRAI
Sbjct: 241  ARIQQNLDQLREELSRAFAGTGTEEGAGPDPQMLSEEVRQRLQAFRQDTYLQIAAFTRAI 300

Query: 918  DQETEEVQQQLAPPPPGHSAFAPEFGQMDSDKALSKLQARLDDLWEDITYSLHDQGHSHL 1097
            DQETEEVQQQLAPPPPGHSAFAPEF Q DS K LSKLQARLDDLWEDIT+SLHDQGHSHL
Sbjct: 301  DQETEEVQQQLAPPPPGHSAFAPEFQQTDSGKVLSKLQARLDDLWEDITHSLHDQGHSHL 360

Query: 1098 GEP 1106
            G+P
Sbjct: 361  GDP 363
```

aminoacid change
with a chemically different aminoacid

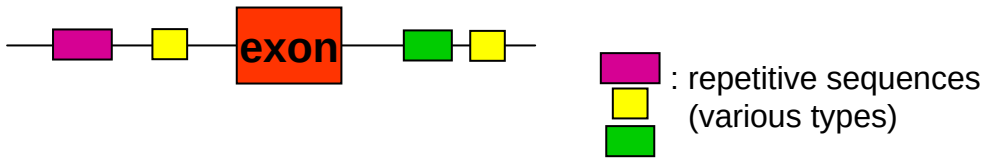
conserved
aminoacid

aminoacid change
with a chemically similar aminoacid
(for example, negatively charged
with negatively charged)

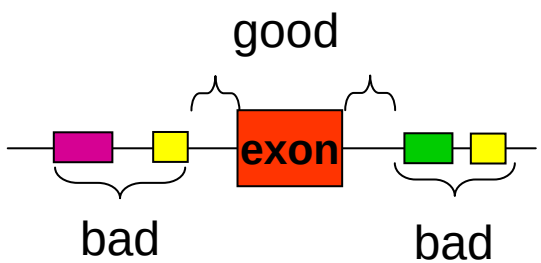
Exercise #8-11: design PCR primer for resequencing exons

Primers to amplify exons for resequencing of clinical samples

- Primers need to be designed in the intronic sequences surrounding the target exon.
- The goal is to pick primers that will only amplify the target region and not other parts of the genome.
- This is largely achieved by avoiding picking primers in repetitive regions of the genome.
- Because of poor sequencing quality near priming sites, it is good practice to design primers at least 30-50 bp away from the exon.

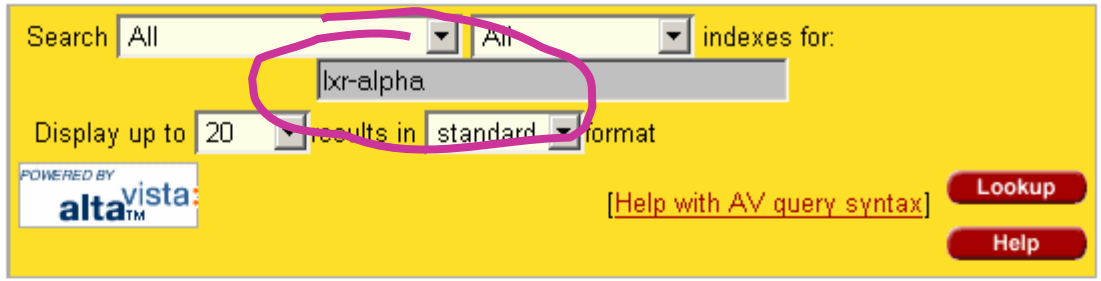


primer design:



Exercise #8: **retrieve the sequence of LXR-alpha exon 4 and of its 2 surrounding introns.**

Use the ENSEMBL genome browser: www.ensembl.org/



1 matches in the *Homo sapiens* Gene index [first 5 matches shown]:

1. Ensembl Gene: ENSG00000025434
Ensembl gene ENSG00000025434 has 2 transcripts: ENST00000344715, ENST00000298843
Oxysterols receptor LXR-alpha (Liver X receptor alpha) (Nuclear orphan receptor LXR-alpha). [Source:Uniprot/SWISSPROT;Acc:Q13133]
The gene has the following external identifiers mapped to it:
GO: GO:0003707, GO:0004887, GO:0005634, GO:0003713, GO:0006355, GO:0003700
HUGO: NR1H3, 7966
LocusLink: 10062
MIM: 602423
protein_id: AAH08819.1, AAA85856.1
RefSeq: NP_005684, NM_005693
Uniprot/SWISSPROT: NRH3_HUMAN, Q13133
http://www.ensembl.org:80/Homo_sapiens/geneview?gene=ENSG00000025434

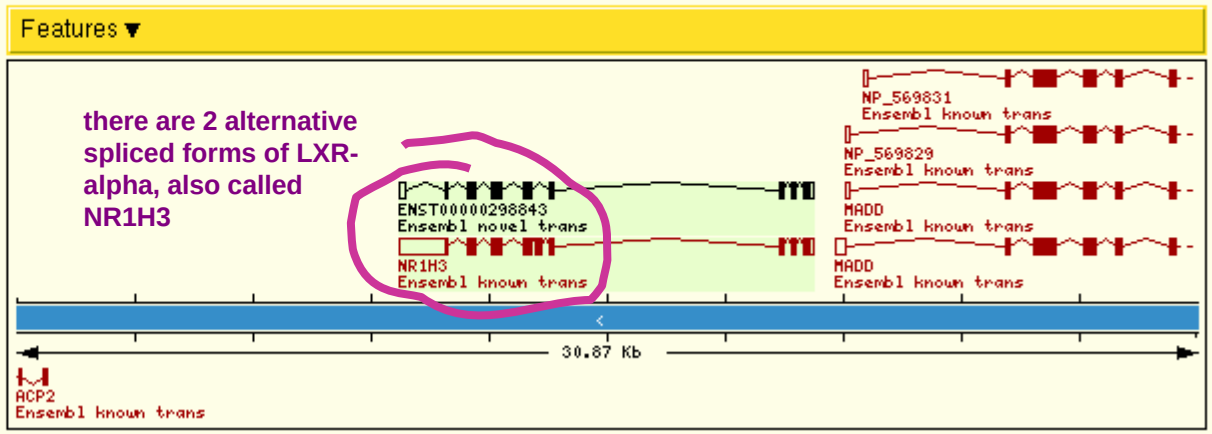
1 matches in the *Mus musculus* Gene index [first 5 matches shown]:

1. Ensembl Gene: ENSMUSG00000002108
Ensembl gene ENSMUSG00000002108 has 1 transcript: ENSMUST00000002177
Oxysterols receptor LXR-alpha (Liver X receptor alpha) (Nuclear orphan receptor LXR-alpha). [Source:Uniprot/SWISSPROT;Acc:Q9Z0Y9]
The gene has the following external identifiers mapped to it:
EMBL: AJ132601, AJ132600, AJ132599, AF085745
LocusLink: 22259

Exercise #8: retrieve the sequence of LXR-alpha exon 4 and of its 2 surrounding introns.

Transcript Structure

1: ENST00000298843 [Transcript information] [Exon information] [Protein information]
2: NR1H3 (ENST00000344715) [Transcript information] [Exon information] [Protein information]



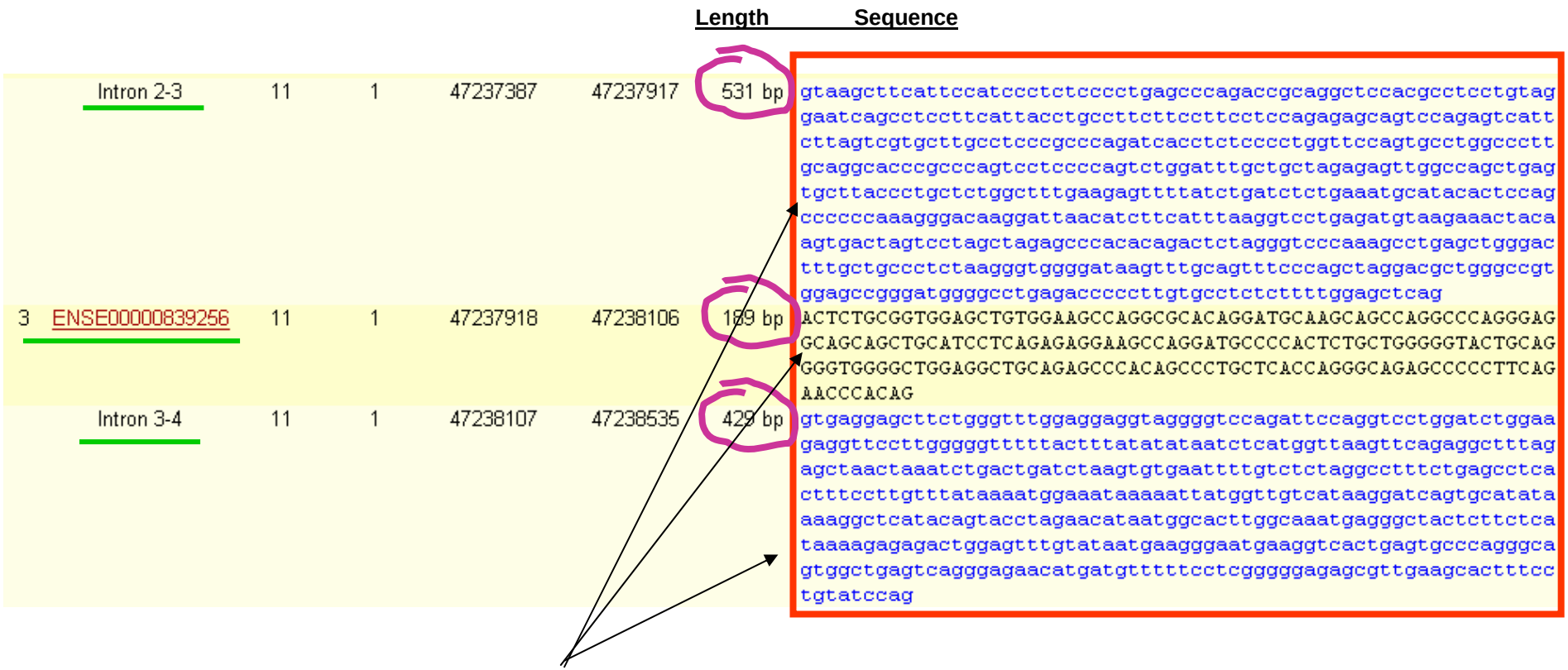
	Intron 7-8	11	1	47246154	47246387	234 bp	gtgtggaggaggggcaatgggaaac.....cctcaactctcttggt
8	ENSE00000713314	11	1	47246388	47246482	95 bp	ACCGGCCCAACGTGCAGGACCAGCTCCAGGTAGAGAGGCTGCAGCACACAT CCCTGCATGCCTACGTCTCCATCCACCATCCCCAT
	Intron 8-9	11	1	47246483	47246676	194 bp	gtgagtctccccatggtgttcctttt.....gtgtttgtctctctcc
9	ENSE00001215906	11	1	47246677	47246965	289 bp	GACCGACTGATGTTCCCACGGATGCTAATGAAACTGGTGAGCCTCCGGACC GTCCACTCAGAGCAAGTGTTTGCACTGCGTCTGCAGGACAAAAAGCTCCCA TCTGAGATCTGGGATGTGCACGAATGACTGTTCTGTCCCCATATTTTCTGT CGGATGGCTGAGGCCTGGTGGCTGCCTCCTAGAAGTGGAACAGACTGAGAA TTCCTGGGAGCTGGGCAAGGAGATCCTCCCGTGGCATTAAAAGAGAGTC
	3' downstream sequence						aaagggttgcgagttttgtggctactgagcagtggagccctcgctaacac.

KEY: -Up/downstream region
-Intron sequence
-UTR region

Display 25 bases either side of intron. View full intron sequence ☑
Display 50 bases of 5' 3' flanking sequence. View only exon sequence ☐
Redraw

Select "Redraw"

Exercise #8: **retrieve the sequence of LXR-alpha exon 3 and of its 2 surrounding introns.**

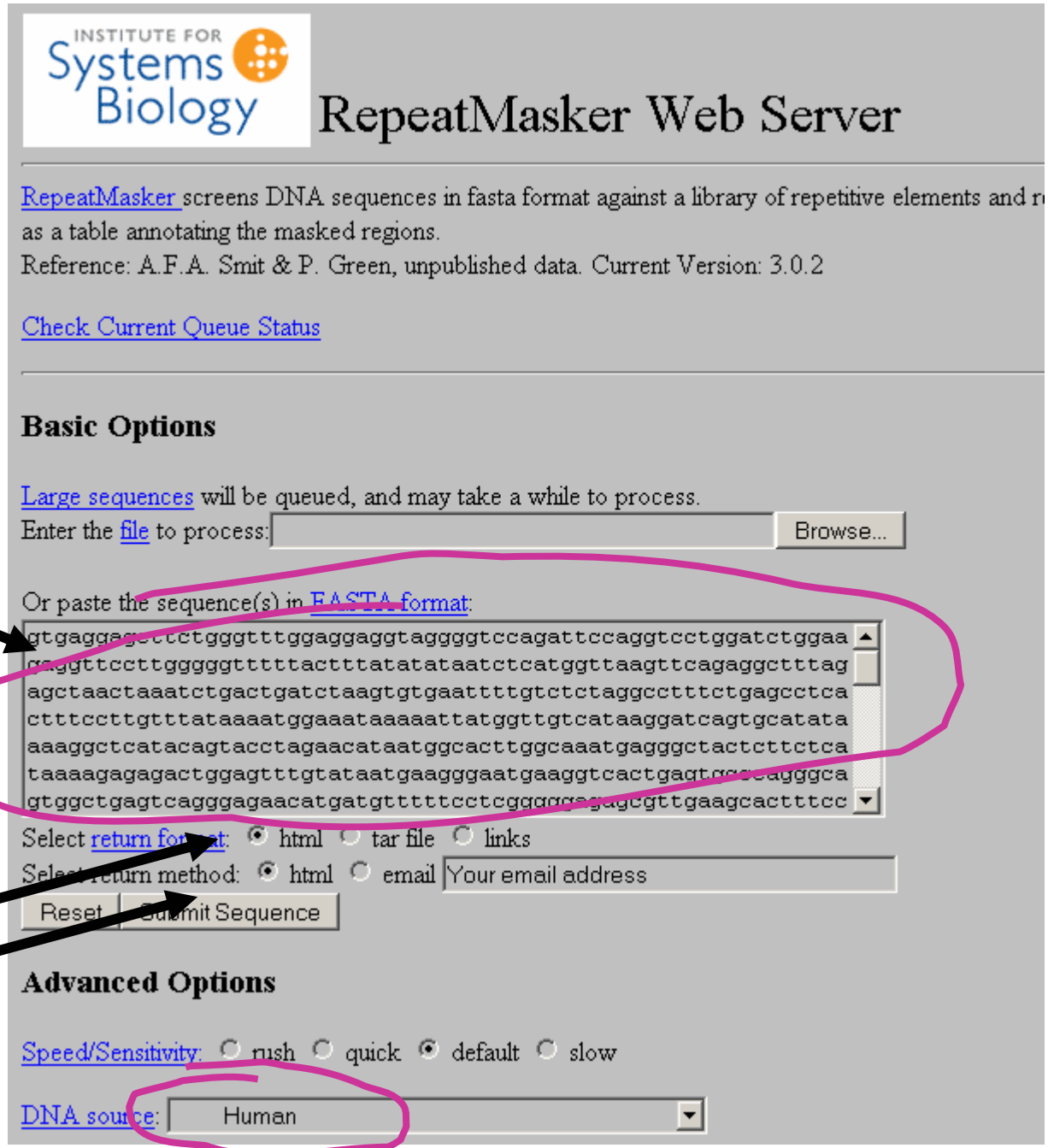


copy these sequences to 1 text file and save on your desktop

Exercise #9: use RepeatMasker www.repeatmasker.org/cgi-bin/WEBRepeatMasker to identify repeats in your sequence

your sequence

output options



Exercise #9: **RepeatMasker results for human LXR-alpha intron 2/3, exon 3 and intron 3/4**

Repeat Annotations:

SW score	perc div.	perc del.	perc ins.	query sequence	position in query begin	end	(left)		matching repeat repeat	class/family	position in repeat begin	end	(left)	ID
377	29.1	15.7	0.0	UnnamedSequence	157	290	(971)	+	MIRb	SINE/MIR	106	260	(8)	1
2300	9.6	0.7	0.0	UnnamedSequence	824	1115	(146)	+	AluSx	SINE/Alu	1	294	(18)	2

position of the repeat within the sequence

repeat type

Masked Sequence:

```
>UnnamedSequence
GTGAGGAGCTTCTGGGTTTGGAGGAGGTAGGGGTCCAGATTCCAGGTCCT
GGATCTGGAAGAGGTTCCTTGGGGGTTTTTACTTTATATATAATCTCATG
GTTAAGTTCAGAGGCTTTAGAGCTAACTAAATCTGACTGATCTAAGTGTG
AATTTNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNCTCTTCTCA
TAAAAGAGAGACTGGAGTTTGTATAATGAAGGGAATGAAGGTCACTGAGT
GCCCAGGGCAGTGGCTGAGTCAGGGAGAACATGATGTTTTTCCTCGGGGG
AGAGCGTTGAAGCACTTTCCTGTATCCAAGATCCGTCCACAAAAGCGGA
AAAAGGGGCCAGCCCCCAAAATGCTGGGGAACGAGCTATGCAGCGTGTGT
GGGGACAAGGCCTCGGGCTTCCACTACAATGTTCTGAGCTGCGAGGGCTG
CAAGGGATTCTTCCGCCGCAGCGTCATCAAGGGAGCGCACTACATCTGCC
ACAGTGGCGGCCACTGCCCCATGGACACCTACATGCGTCGCAAGTGCCAG
GAGTGTCGGCTTCGCAAATGCCGTCAGGCTGGCATGCGGGAGGAGTGTGA
GTTTCTGGGGCTGGAGTGGGGAAGAGGCTGAGGGGAAAGAGGGGGCCAGG
GTGTGACCCAAAACAGGTGCCTGAACTTGCAGGGGCTAACTGATCCCTAA
GTATGGATCCCAGTATCTTTCTTNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNTCTTCTTGCCTTTACCCAGTGCTGTCTGCTTTTCT
GGAGCCCCAAACCACCCCCTTTGCCCCATCCTTCCCTCCTGTCTTTCCCC
CACCCCCTTGCCCCATCCTTTCCCCATCTGCTCCCTTCCTCATATTTGGC
CCTGTCCTTAG
```

exon sequence

masked repeat

Exercise #10: Use Primer3 frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi to design primers to amplify your exon-containing sequence

Primer3

pick primers from a DNA sequence

disclaimer

cautions

Paste source sequence below (5'->3', string of ACGTNacgtn -- other letters treated as N -- numbers and blanks ignored). FASTA sequence (vector, ALUs, LINEs, etc.) or use a Mispriming Library (repeat library): NONE

```
GCCCAGGGCAGTGGCTGAGTCAGGGAGAACATGATGTTTTTCCTCGGGGG
AGAGCGTTGAAGCACTTTCCTGTATCCAG[AGATCCGTCCACAAAAGCGGA
AAAAGGGGCCAGCCCCCAAAATGCTGGGGAACGAGCTATGCAGCGTGTGT
GGGGACAAGGCCTCGGGCTTCCACTACAATGTTCTGAGCTGCGAGGGCTG
CAAGGGATTCTTCCGCCGCAGCGTCATCAAGGGAGCGCACTACATCTGCC
ACAGTGGCGGCCACTGCCCCATGGACACCTACATGCGTCGCAAGTGCCAG
GAGTG]TCGGCTTCGCAAATGCCGTCAGGCTGGCATGCGGGAGGAGTGTGA
```

you can mark the boundary of the exon junction in the sequence with [ ]
for exon resequencing, leave 50 extra bases on both sides

☑ Pick left primer or use left primer below.

☐ Pick hybridization probe (internal oligo) or use oligo below.

☑ Pick right primer or use right primer below (5'->3' on opposite strand).

General Primer Picking Conditions

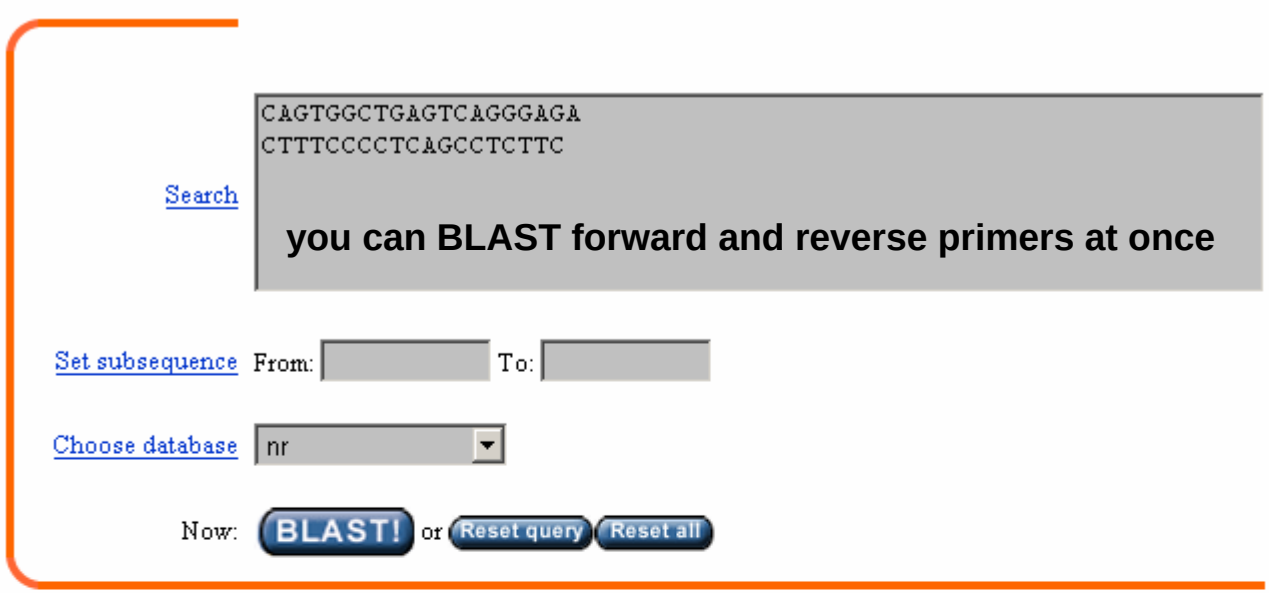
Primer Size	Min: 18	Opt: 20	Max: 27				
Primer Tm	Min: 57.0	Opt: 60.0	Max: 63.0	Max Tm Difference: 100.0			
Product Tm	Min:	Opt:	Max:				
Primer GC%	Min: 20.0	Opt:	Max: 80.0				
Max Self Complementarity: 8.00				Max 3' Self Complementarity: 3.00			
Max #N's: 0				Max Poly-X: 5			
Inside Target Penalty:				Outside Target Penalty: 0		Set Inside Target Penalty t	
First Base Index: 1				CG Clamp: 0			
Salt Concentration: 50.0				Annealing Oligo Concentration: 50.0		(Not the concentration of	

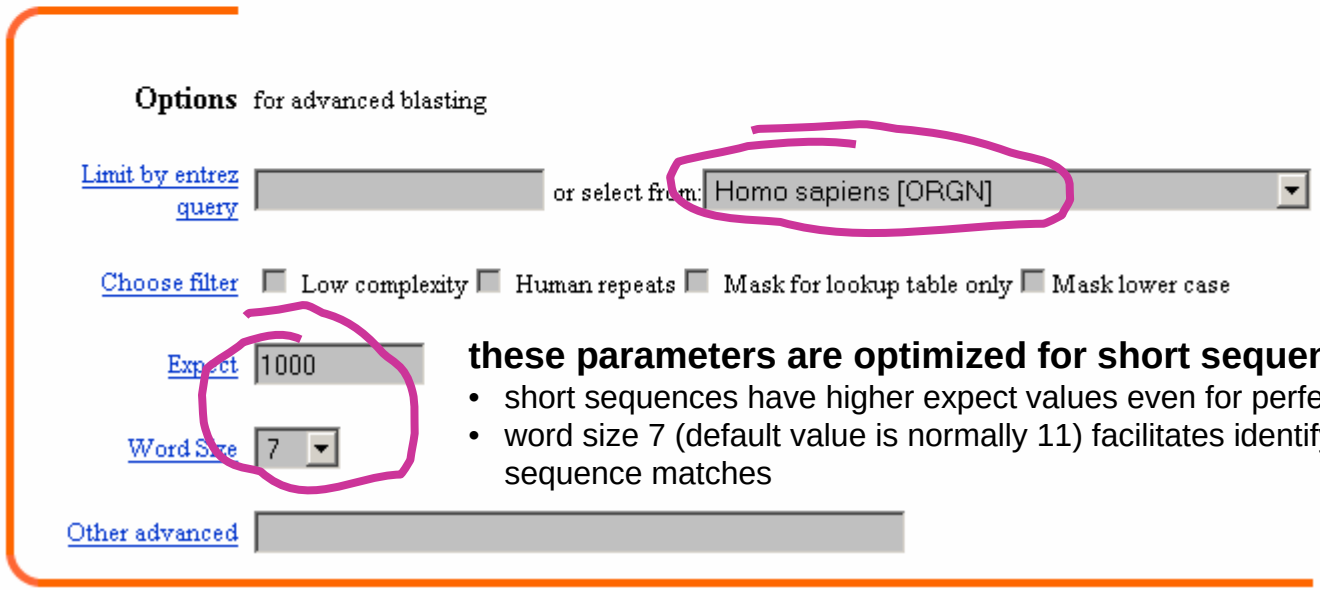
☑ Liberal Base ☐ Show Debuging Info ☑ Do not treat ambiguity codes in libraries as consensus

Pick Primers | Reset Form

default parameters are normally good

Exercise #11: **Use BLAST for short, nearly exact matches (see BLAST page) to verify that your primers are unique in the human genome**





these parameters are optimized for short sequence matches:

- short sequences have higher expect values even for perfect matches
- word size 7 (default value is normally 11) facilitates identifying short sequence matches

Exercise #11: **Inspect BLAST output: you want to see that your primers only hit your target region**

```
                                                                    Score     E
Sequences producing significant alignments:                         (bits) Value

gi|21622769|gb|AC090589.9|  Homo sapiens chromosome 11, clon...     40    0.071
gi|24850147|gb|AC018410.24|  Homo sapiens chromosome 11, clo...     40    0.071
gi|29124047|gb|AC010427.5|  Homo sapiens chromosome 5 clone ...     38    0.28
gi|28827851|gb|AC026740.6|  Homo sapiens chromosome 5 clone ...     38    0.28
gi|21955075|gb|AC105935.2|  Homo sapiens chromosome 3 clone ...     36    1.1
```

correct chromosome

```
>gi|21622769|gb|AC090589.9|  Homo sapiens chromosome 11, clone RP11-390K5, complete sequence
          Length = 190017

 Score = 40.1 bits (20), Expect = 0.071
 Identities = 20/20 (100%)
 Strand = Plus / Plus

Query: 1      cagtggctgagtcagggaga 20
               ||||||||||||||||||||
Sbjct: 150797 cagtggctgagtcagggaga 150816
```

```
 Score = 38.2 bits (19), Expect = 0.28
 Identities = 19/19 (100%)
 Strand = Plus / Minus

Query: 21     ctttcccctcagcctcttc 39
               |||||||||||||||||||
Sbjct: 151177 ctttcccctcagcctcttc 151159
```

wrong chromosome

```
>gi|29124047|gb|AC010427.5|  Homo sapiens chromosome 5 clone CTD-2198K18, complete sequence
          Length = 68883

 Score = 38.2 bits (19), Expect = 0.28
 Identities = 19/19 (100%)
 Strand = Plus / Minus

Query: 16   ggagactttcccctcagcc 34
             ||||||||||||||||||
Sbjct: 8525 ggagactttcccctcagcc 8507
```

this match is a mix of the forward and reverse primer and can be ignored